It Takes Two to Tango: DNA Polymerase Responses to Replication Stress in Human Cells

Kristin A. Eckert, Ph.D.
Professor of Pathology
Gittlen Cancer Research Foundation
Penn State University, College of Medicine

Normal DNA

Microsatellites

National Institute on Aging
Laboratory of Molecular Gerontology
January 18, 2016

The Human Genome

67% of the genome is repetitive sequence

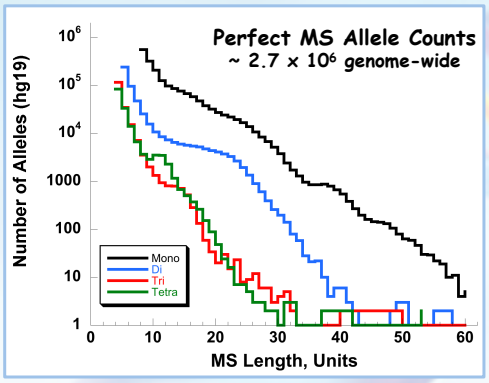
Repetitive sequences impact mutagenesis and genome evolution

Normal DNA

Microsatellites

The Human Genome

67% of the genome is repetitive sequence



Microsatellite Sequences

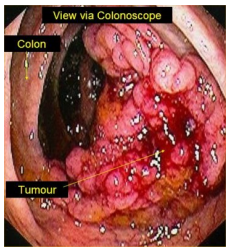
- Interspersed tandem repeats
- Repeat motif = 1-6 basepairs
- Genic and intergenic regions

How are repetitive sequences accurately and efficiently replicated in human cells?

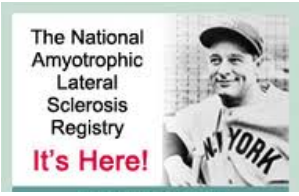
Microsatellites are Functional

Dog breed skeletal morphology (Runx-2 gene polymorphisms)

FMR-1 Repeat expansions



Colorectal cancer (MSI)



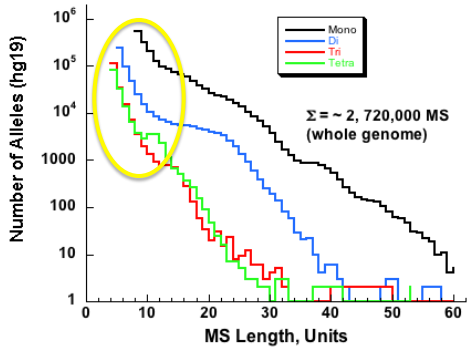
ATXN2 (SCA2) Interruptions

Genic Microsatellite Length Polymorphisms Affect Disease Risk

Gene	Allele	Effect	Disease	Reference
EGFR	$[CA]_{14-22}$ Intron 1	Transcription efficiency	Breast cancer	Gebhardt 1999 Buerger 2000 Buerger 2004
eNOS	$[CA]_{17-44}$ Intron 13	Splicing efficiency	Coronary artery disease	Stang 2000 Hui 2003; 2005
HO-1	$[CA]_{20-30}$	Promoter efficiency	Coronary artery disease	Chen 2002
CFTR	$[GT]_{9-13}$; $[T]_n$ Exon 9;	Splicing efficiency	Cystic fibrosis severity	Hefferson 2004
UGT1A1	$[TA]_{5-8}$; promoter	Transcriptional activity	Head/heck cancer; Gilbert's syndrome	Bosma 1995 Lacko 2010
IL28A	$[TA]_{10-18}$; promoter	Transcriptional activity	Response to IFN therapy	Sugiyama et al 2011

Lengths of microsatellite repeats directly affect mRNA production

How are microsatellite sequences accurately replicated in human cells?



~ 1.6 million genic microsatellites
(5' and 3' UTRs, promoters, introns, exons)

M.S. Hershey Medical Center

Millenium Science Building

PENNSTATE 1855

Eckert Laboratory
Penn State-Hershey
College of Medicine

Makova Laboratory
Penn State-University Park
College of Science

Experimental
- Biochemical
- Genetic

Computational
- Genomics
- Statistical models

What is a Microsatellite ?

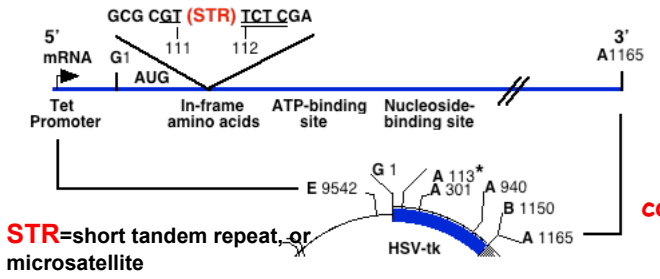
At what length does a tandem repeat become a microsatellite?

Definition based on mutational behavior

-combined computational and experimental approaches

Kelkar et al *Genome Biology and Evolution*. 2010; 2: 620-635.
Ananda, Walsh et al *Genome Biology and Evolution*, 2013; **5**: 606-620.

Microsatellite Mutagenesis – HSV-tk Forward Assay



STR=short tandem repeat, or microsatellite

STR alleles reflect common sequences and lengths present in human genome

Mutations detectable:

- Base substitutions
- Frameshifts (Indels)
- Deletions up to 5kB
- Rearrangements
- Microsatellite instability

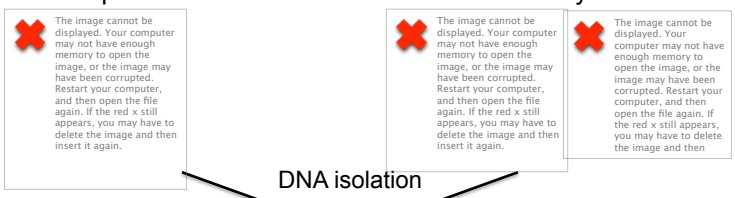
Mono	Di	Tetra
A/T	GT/CA	TTCC/AAGG
G/C	TC/AG	TTTC/AAAG
	AT/TA	TCTA/AGAT

Hile et al (2000) Cancer Research 60: 1698.

Overview of the HSV-tk Mutagenesis System

Genetic
Ex vivo (human cells)
DNA Replication

Biochemical
In vitro (Human proteins)
DNA synthesis



DNA isolation

HSV-tk Mutant Frequency (Quantitative)

↓

Mutational Specificity (Qualitative)

Definition of a Microsatellite

Mature Microsatellite

GTGTGTGTGT ⟷ GTGTGTGTGTGT

CACACACACA

GTGTGTGTGTGTGTGT

$[GT/CA]_n$

GTGTGTGTGTGTGTGTGTGT

Polymorphic allele

mutability

Threshold lengths
Mono: 9 units
Di: 5 units
Tri: 4 units
Tetra: 4 units

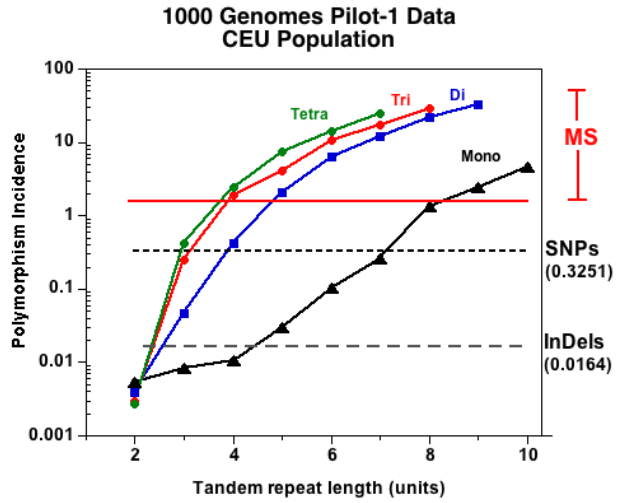
↑ "birth"

GTGTGTTTGT
CACACAAACA

Short tandem repeat

Kelkar et al *Genome Biology and Evolution*. 2010; 2: 620-635.
Ananda, Walsh et al *Genome Biology and Evolution*, 2013; **5**: 606-620.

Microsatellites are Highly Polymorphic in Human Populations



- Polymorphism incidence used as a proxy for microsatellite mutability

G. Ananda, E. Walsh et al., Genome Biol. Evol. (2013) 5(3):606-20

Microsatellites are Stable in Mitotic Human Cells

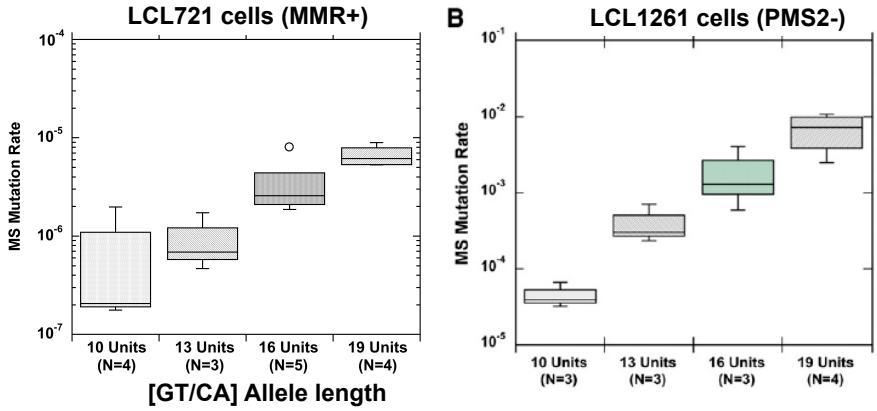
Variable

	Motif Sequence	Mutation Frequency per cell generation x 10^{-6}	
		Median	Range
motif size	$[G/C]_{10}$	29	14-34
	$[GT/CA]_{10}$	**0.21**	0.18-2.0
	$[GT/CA]_{16}$	3.4	2.1-11
Number (length)	$[TC/AG]_{11}$	2.3	0.59-7.5
	$[TC/AG]_{17}$	9.8	6.0-28
	$[TC/AG]_{20}$	21	16-25
sequence	$[TTCC/AAGG]_9$	5.6	3.8-20
	$[TTTC/AAAG]_9$	35	12-57
	$[TCTA/AGAT]_9$	**48**	36-110

Measurements performed using LCL721 cell line, derived from a normal donor

Eckert et al (2002); Eckert and Hile (2009)

Mature Microsatellite Mutation Rates



- **Mismatch repair stabilizes microsatellites, but also introduces mutational biases**

Baptiste et al (2013) G3: Genes, Genomes, Genetics; Shah and Eckert (2009) Cancer Research

It takes a village to raise a child….

Multiple DNA polymerases are required to maintain human genome stability

Replication

Pol alpha (α, *POLA*)
Pol delta (δ, *POLD1*)
Pol epsilon (ε, *POLE*)

Repair

Pol beta (β, *POLB*)
Pol lambda (λ, *POLL*)
Pol mu (μ, *POLM*)
Pol theta (θ, *POLQ*)

Specialized Synthesis

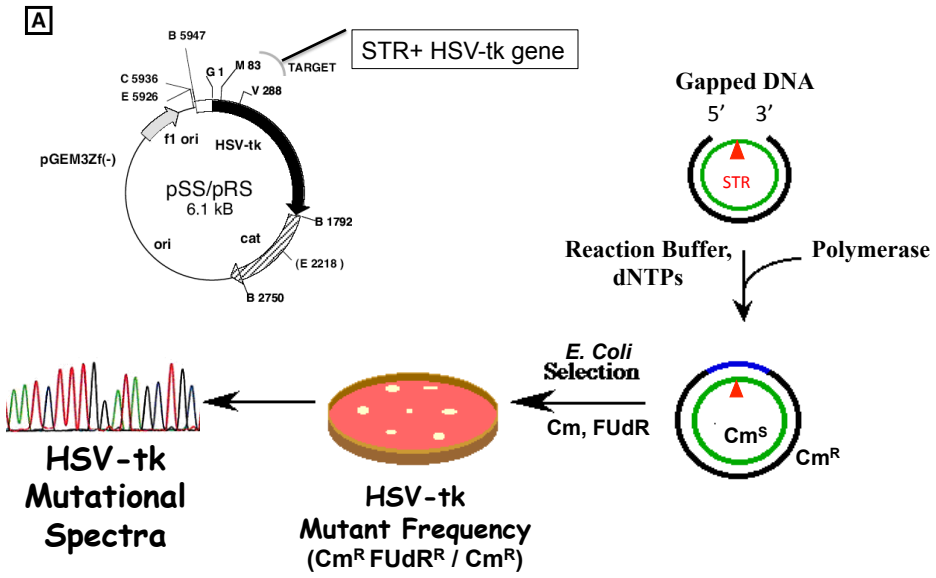
Pol zeta (ζ, *POLZ; REV3*)
Pol eta (η, *POLH; XPV*)
Pol kappa (κ, *POLK; DINB1*)
Pol iota (ι, *POLI*)
Rev 1 (dCMP transferase)
Pol nu (ν, *POL N*)

Mitochondrial

Pol gamma (γ, *POLG*)

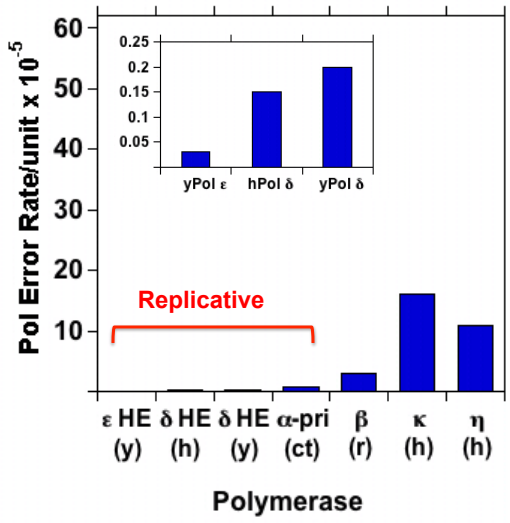
Are specialized polymerases required for accurate replication of repetitive DNA ?

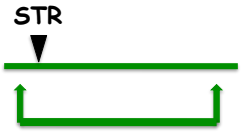
In vitro HSV-tk Polymerase Assay



Eckert, Mowery and Hile (2002) Biochemistry 41: 10490-98.

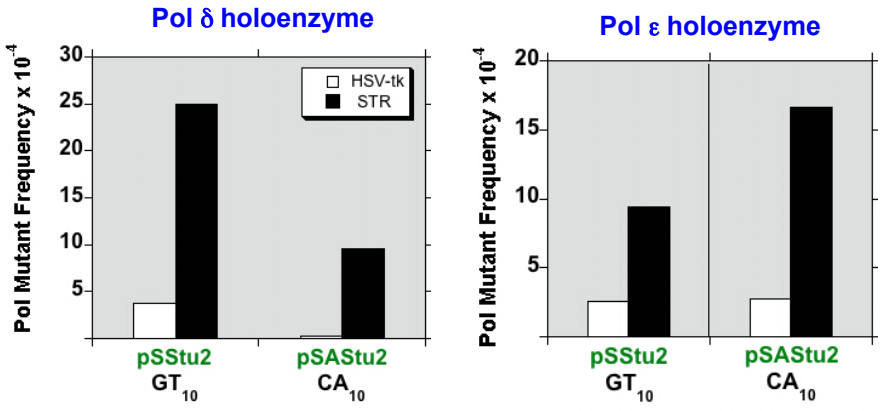
Replicative Polymerases Display High Accuracy within the HSV-tk Gene





HSV-tk Sequences (frameshift monitors)

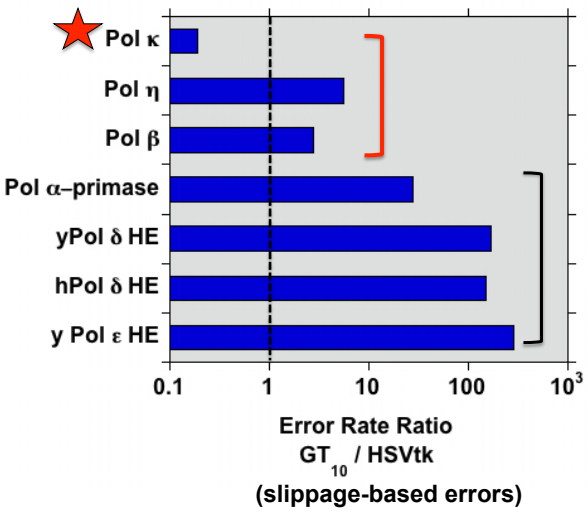
Replicative holoenzymes are highly inaccurate within microsatellites

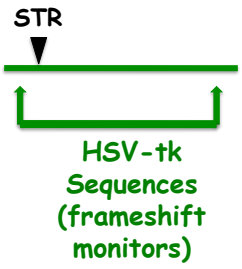


- The 3' to 5' proofreading exonuclease does <u>not</u> remove microsatellite indel errors

Abdulovic, Hile, Kunkel and Eckert (2011) DNA Repair 10: 497.

Specialized polymerases are more accurate for motif-based InDel errors





Hile et al. (2011) Nucl. Acids Res. 40: 1636

Microsatellite Mutational Specificity

$[GT/CA]_9 \longleftarrow [GT/CA]_{10} \longrightarrow [GT/CA]_{11}$

Base Deletion Error

$[GT]_4\ T\ [GT]_5$

Base Substitution Error

$[GT]_4\ TT\ [GT]_5$

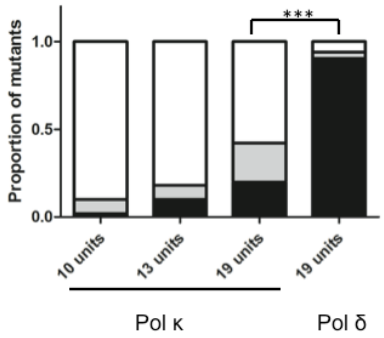
Base Insertion Error

$[GT]_5\ G\ [GT]_5$

Microsatellite Interruptions

Pol κ Displays a Unique Error Specificity within Microsatellites

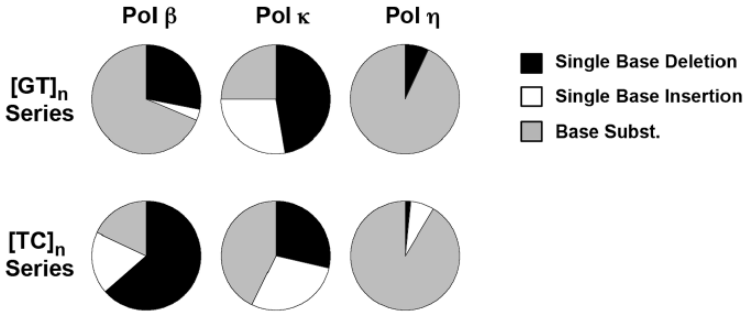
White bars: HSVtk gene errors
Black bars: Microsatellite errors
*** $p=2.9 \times 10^{-4}$ (Fisher's)



Observed for di- and tetranucleotide microsatellites

Hile et al *Nucleic Acids Res.* 2012; 40: 1636-47
Baptiste and Eckert *Environ. Mol. Mutagenesis* 2012; **53:** 787-796.

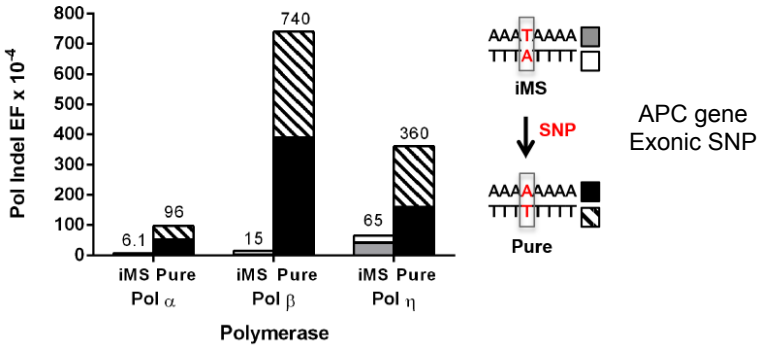
Specialized DNA Polymerases Create Interruption Errors



The specificity of in vitro polymerase errors is in concordance with genome-wide analyses of interrupted microsatellites within individual human genomes

Ananda et al. *PLoS Genetics*, 2014; 10(7):e1004498.

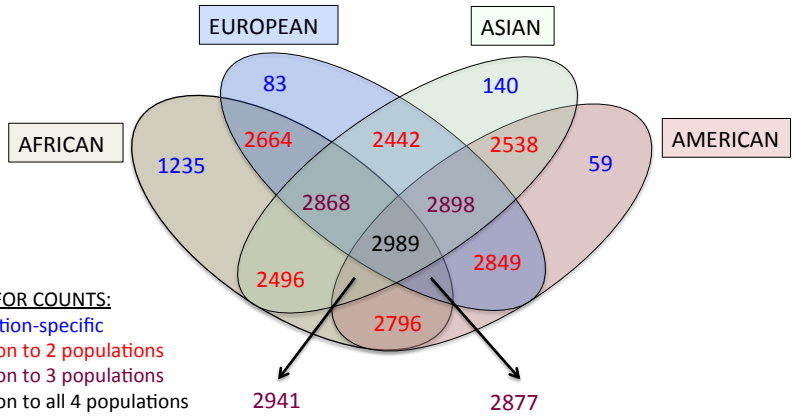
Interruptions Decrease Polymerase Errors



Interrupting base is rarely deleted

Ananda et al. *PLoS Genetics*, 2014; 10(7):e1004498.

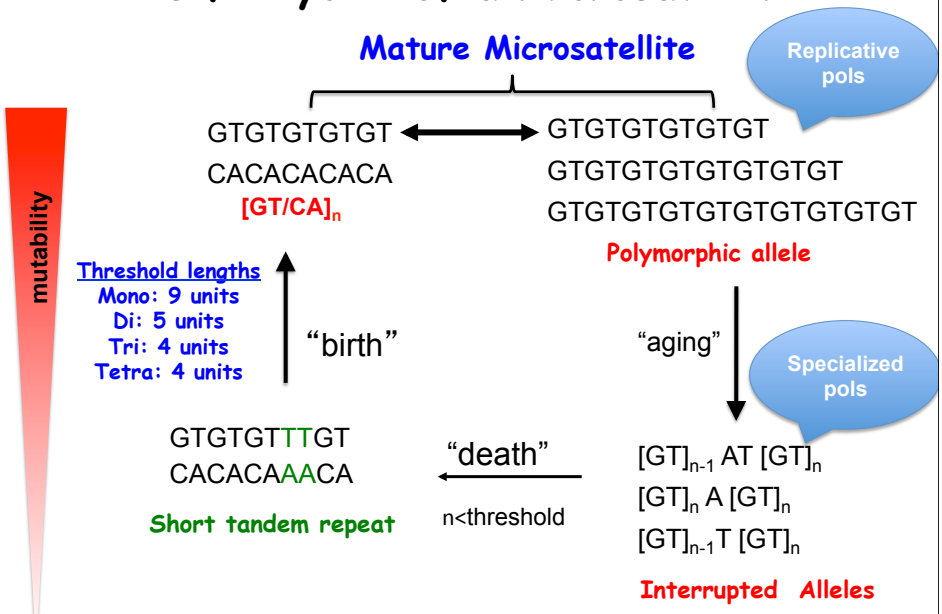
Microsatellite interruptions occur within individual genomes



COLORS FOR COUNTS:
1. Population-specific
2. Common to 2 populations
3. Common to 3 populations
4. Common to all 4 populations

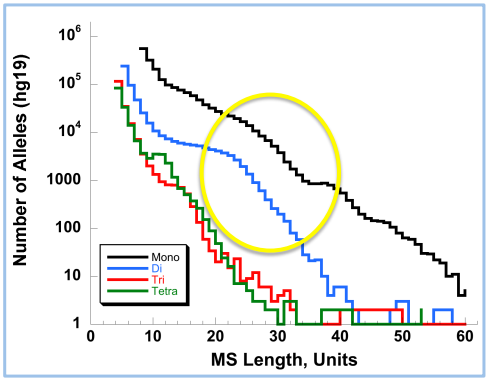
~11% of interrupted microsatellites lie within exons (UTRs + coding)

"Life Cycle" of a Microsatellite



Ananda et al. *PLoS Genetics*, 2014; 10(7):e1004498.

How are microsatellite sequences efficiently replicated in human cells?



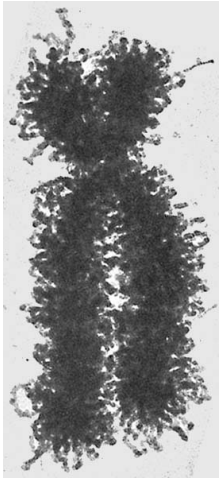
Chromosome Fragile Sites

Telomere

Rare fragile sites

Common fragile sites (CFS)*

Interstitial Telomere



$[TTAGGG]_n$

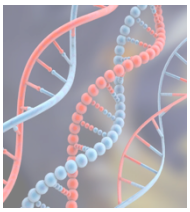
$[CGG]_n$

$[A/T]_n$
$[AT/TA]_n$

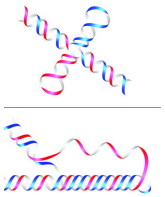
Microsatellite DNA

* Low complexity A/T-rich sequences and mononucleotide A/T microsatellites are significant predictors of aphidicolin-induced CFS genome-wide.
Fungtammasan et al. (2012) Genome Research 22: 993-1005.

Repetitive Sequences Favor Non-B Form DNA Secondary Structures



B-DNA ⟷ **Non B-DNA**



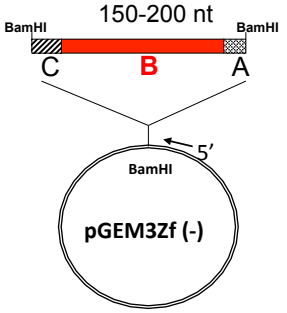
B-DNA structure is favored in random sequences

Non-B DNA Structures

- Slipped strands
- Bent DNA $(A)_n$ $(T)_n$
- Triplex (H) $(TC)_n$ $(AAG)_n$
- Cruciforms $(AT)_n$
- Z-DNA $(GT)_n$ $(GC)_n$
- Hairpins $(AT)_n$ $(CAG)_n$
- G4 $(TTAGGG)_n$

In Vitro Assay for DNA Polymerase Pausing

Template	CFS	Repeat Elements
AT repeat 1	FRA16D	$[A/T]_{28}$, $[AT/TA]_{23i}$
AT repeat 2	FRA16D	$[A/T]_{9}$, $[AT/TA]_{8}$
AT repeat 3	FRA3B	$[A/T]_{22}$, $[AT/TA]_{25}$
AT repeat 4	FRA16D	$[AT/TA]_{9}$
Quasipalindrome 1	FRA16D	$[A/T]_{19}$, QP_{36}
Quasipalindrome 2	FRA16D	QP_{37}
Quasipalindrome 3	FRA16D	QP_{29}

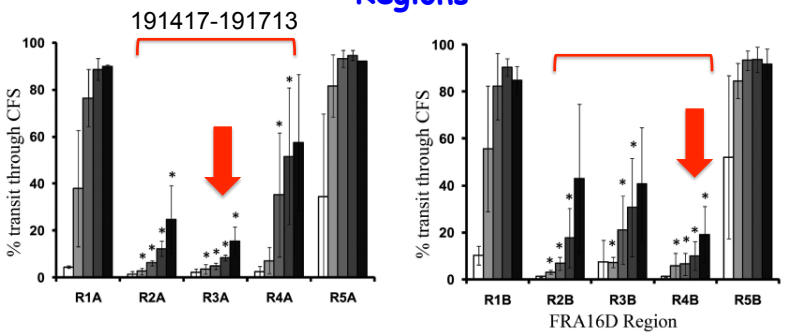


A = constant proximal
B = CFS sequence
C = constant distal

% Transit = ∑ Products C/B+C

Shah et al. (2010) Nucl. Acids Res. 38: 1149-1162.
Walsh et al. (2013) J. Mol. Biol. 425: 232-43.

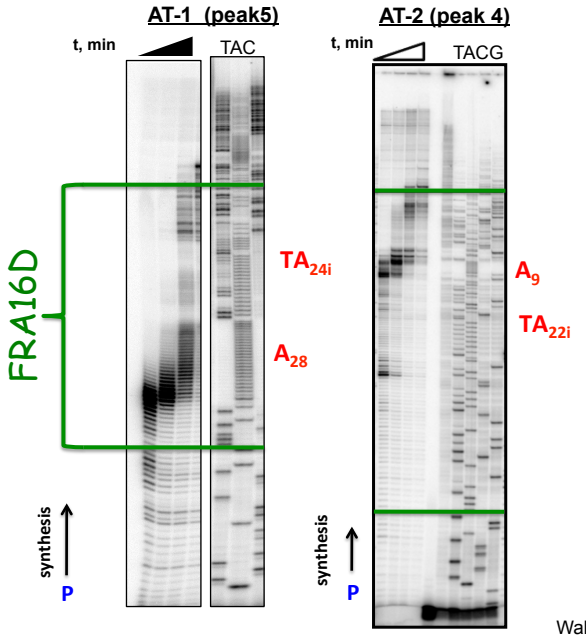
Slowed Elongation by Human Polδ4/PCNA within FRA16D Regions

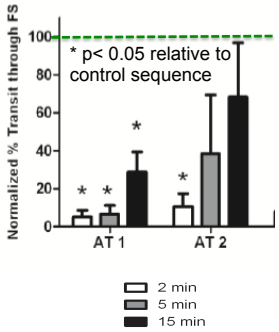


- Slowed DNA synthesis within specific sequences
- Strand biases observed
- Similar results observed in HeLa cell extracts

Shah et al (2010) Nucl. Acids Res.38: 1149.

Pol δ4 Inhibition: A/T and AT/TA Microsatellites

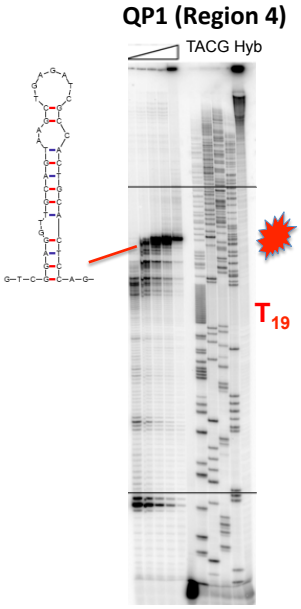




Control= sequence from FRA16D with high %AT but no repeated motifs

Walsh et al. (2013) J. Mol. Biol. 425: 232.

Pol δ4 Inhibition: Quasipalindromes

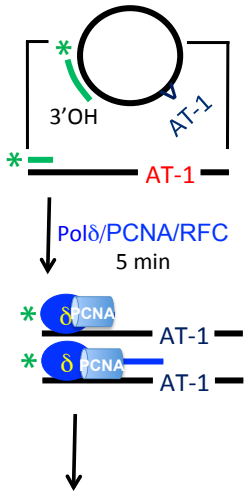


Motif	Stem length (bp)	Stem self-homology	Loop length (nt)	Pol δ TP
QP1	29	76 %	7	0.51
QP2	34	82%	3	0.38
QP3	20	100 %	9	0.72

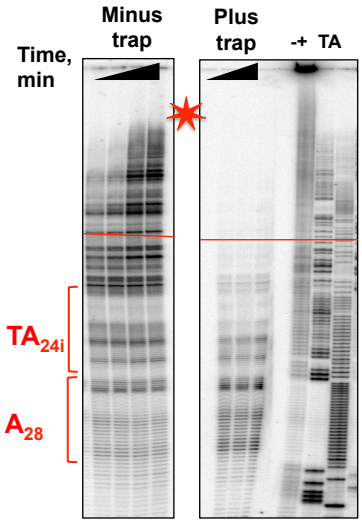
TP = termination probability at base of QP

Walsh et al. (2013) J. Mol. Biol. 425: 232.

hPol δ4/PCNA Holoenzyme Dissociates at Inhibitory Sequences



Add cold ssDNA to trap dissociated pol
Monitor synthesis



Walsh et al. (2013) J. Mol. Biol. 425: 232.

CFS Motifs Inhibitory to Replicative Polymerases

Sequence Motif*	Replicative Polymerase		
	Pol δ4	ε (cat)	α–primase
Microsatellites			
$[A/T]_9$	+/-		
$[A/T]_{19}$	+	+	+
$[A/T]_{22}$	+		
$[A/T]_{28}$	+++	+/-	++
$[AT/TA]_8$	–		
$[AT/TA]_{24i}$	+/-	++	++
$[AT/TA]_{25}$	++		
Inverted repeats 29 nt stem (100%)	+++		
36 nt stem (75%)	+++	+++	++
37 nt stem (82%)	+++		

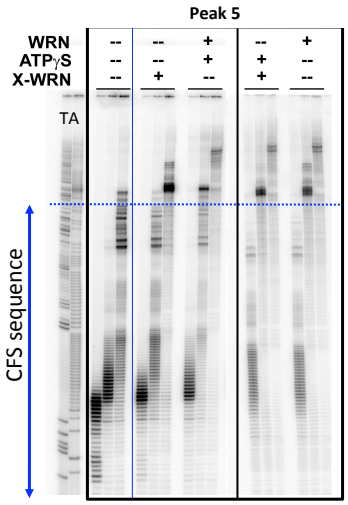
Consistent with genome-wide features

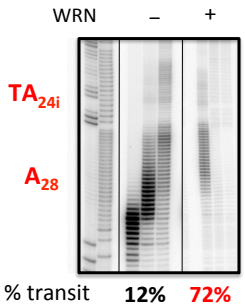
* Strand biases observed

Do specific *trans*-acting replication proteins alter DNA synthesis progression?

Polymerase-interacting proteins (RecQ helicases)

WRN facilitates hPol δ4 progression: Microsatellites

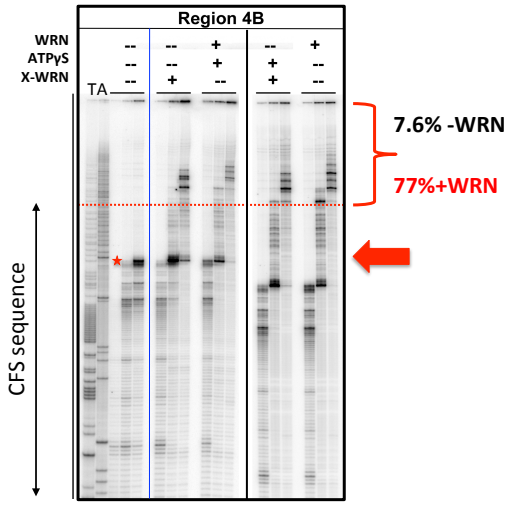




Neither helicase nor exonuclease activities of WRN are necessary for pol δ stimulation

Shah et al (2010) Nucl. Acids Res.38: 1149.

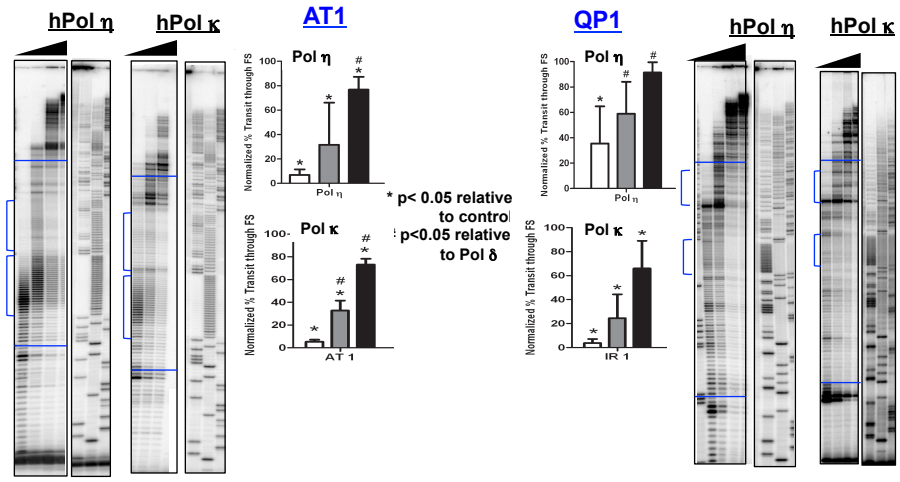
WRN facilitates hPol δ4 progression: Hairpin Structure



- WRN has no effect on pol δ synthesis efficiency through control sequences
- WRN effect due to direct interaction with either Pol δ and/or the DNA template

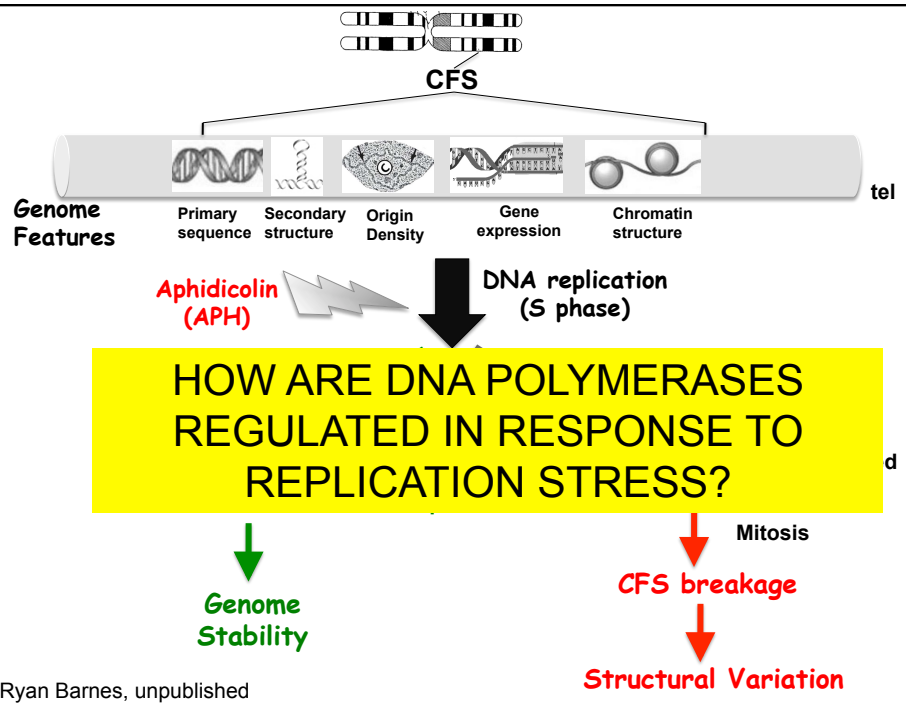
Do specific *trans*-acting replication proteins alter DNA synthesis progression?

Y Family polymerases display efficient synthesis through CFS repeat elements



Common 3 polymerase reaction conditions:
KPO_4 buffer (pH7.6)
80% extension of control template

Bergoglio et al. (2013) J. Cell Biol. 201: 395-408.
Walsh et al. (2013) J. Mol. Biol. 425: 232-243.



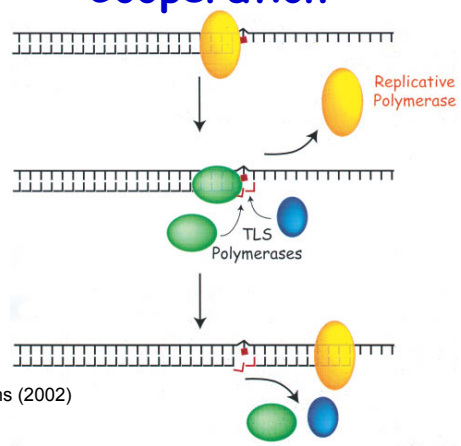
Ryan Barnes, unpublished

Diverse functions of DNA Polymerase η

1. During natural, targeted mutation
   - immune cell somatic hypermutation
2. At DNA lesions during replication (translesion synthesis)
   - UV photoproducts; cis-platinum adducts
3. **Common fragile site replication**

It Takes Two: DNA Polymerase Cooperation

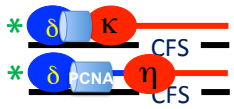
General model



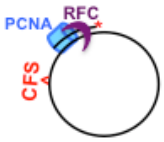
Taken from: Pages and Fuchs (2002)

Does polymerase exchange take place at inhibitory CFS-derived sequences?

In vitro assay to examine DNA polymerase cooperation



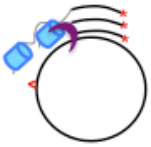
(1) RFC loading of PCNA onto primed ssDNA templates



Add Pol δ4

37°C, 5 min

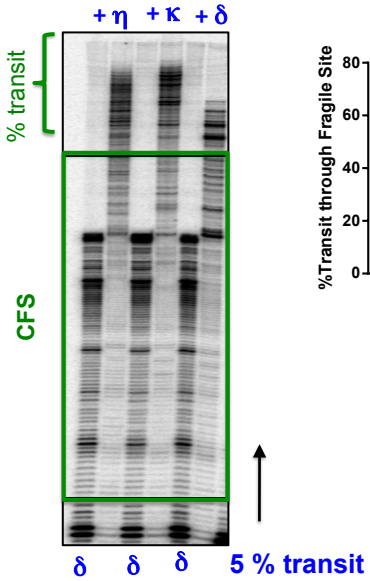
(2) Pol δ HE formation and initial synthesis

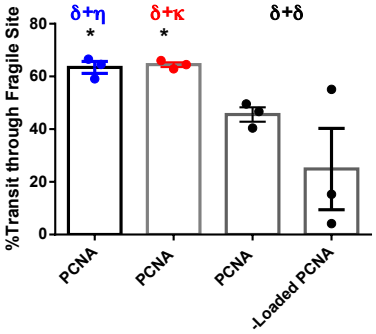


Equal activity
+ Pol η (κ)
-or-
+ Pol δ4 (control)

Monitor synthesis At 15 min (T_2)

Specialized Polymerases Efficiently Exchange with a Moving Pol δ Holoenzyme



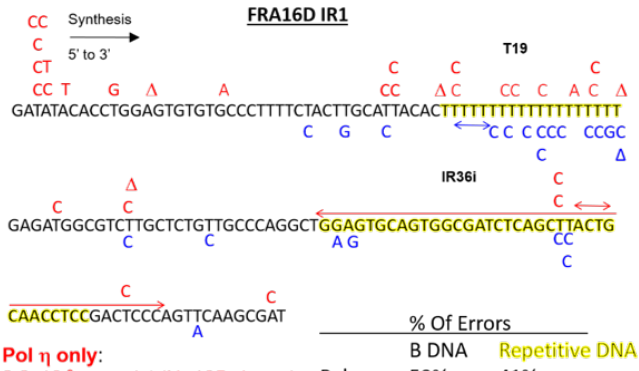


*Significantly different from δ control reactions; P<0.05

N= 3 independent reactions

Mono- Pol η reaction
% transit = 38%

Pol η Error Signatures are Present at CFS Inhibitory Motifs in Dual Polymerase Reactions

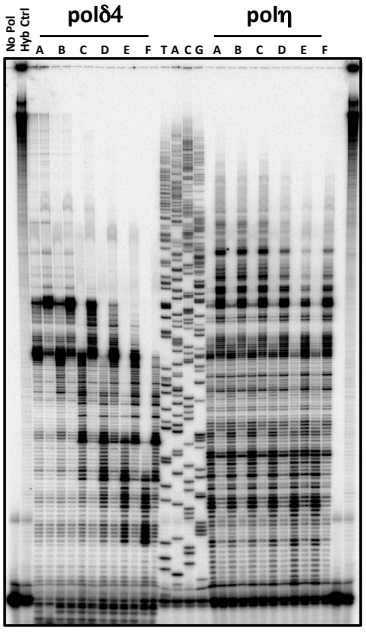


Pol η only:
2.2x10^{-3} errors/nt (N=105 clones)
Pol δ only:
0.1x10^{-3} errors/nt (N=60 clones)
Pols η+δ:
1.7 x 10^{-3} errors/nt (N=94 clones)

	% Of Errors	
	B DNA	Repetitive DNA
Pol η	59%	41%
Pol δ + η	25%	75%

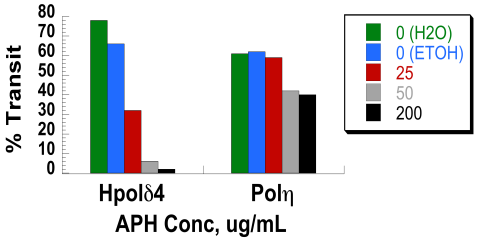
Fishers exact test – P=0.017

Hpolδ4, but not Polη is Inhibited by APH

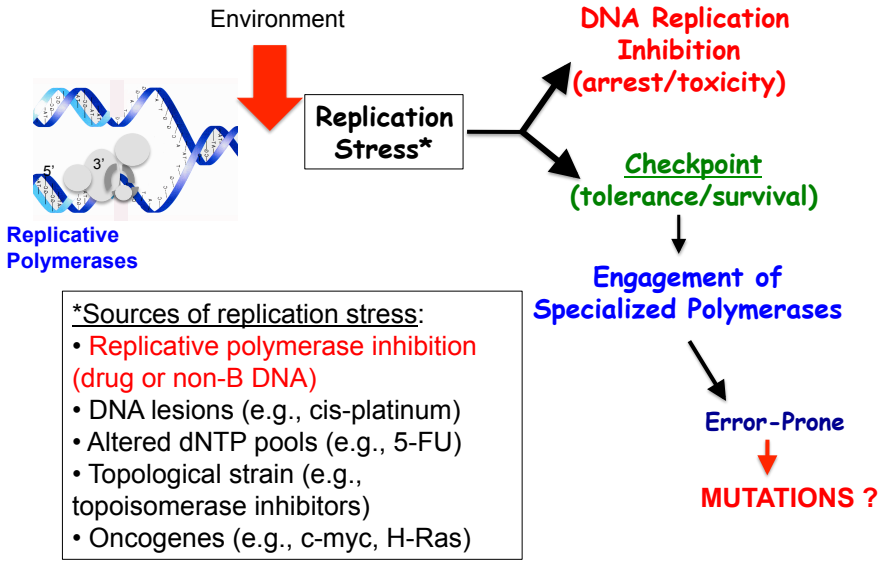


A = H_2O control
B = ETOH control
C = 25ug/mL APH
D = 50ug/mL APH
E = 100ug/mL APH
F = 200ug/mL APH

Template = pGEM 3Zf (-) +PCNA/RFC



Replication Stress and Genome Instability



*Sources of replication stress:
- Replicative polymerase inhibition (drug or non-B DNA)
- DNA lesions (e.g., cis-platinum)
- Altered dNTP pools (e.g., 5-FU)
- Topological strain (e.g., topoisomerase inhibitors)
- Oncogenes (e.g., c-myc, H-Ras)

It takes a village to raise a child….

Multiple DNA polymerases are required to maintain human genome stability

Replication

Pol alpha (α, *POLA*)*
Pol delta (δ, *POLD1*)*
Pol epsilon (ε, *POLE*)*

Repair

Pol beta (β, *POLB*)*
Pol lambda (λ, *POLL*)
Pol mu (μ, *POLM*)
Pol theta (θ, *POLQ*)

Specialized Synthesis

Pol zeta (ζ, *POLZ; REV3*)
Pol eta (η, *POLH; XPV*)*
Pol kappa (κ, *POLK; DINB1*)*
Pol iota (ι, *POLI*)*
Rev 1 (dCMP transferase)
Pol nu (ν, *POL N*)
PrimPol *

Mitochondrial

Pol gamma (γ, *POLG*)

* Studied in the Eckert laboratory

Perspective

The number and wide ranging activities of mammalian DNA polymerases may reflect the biochemical complexity of completing replication of a repeat-dense genome.

The Team

Eckert Laboratory

- Suzanne Hile
- Ryan Barnes
- Bryan Tsao

Former grad students:

- Erin Walsh
- Sandeep Shah

Computational (UP campus)

- Kateryna Makova
- Francesca Chairomonte

Former grad students:

- Chai Fungtammassen
- Guru Ananda

Collaborators/Contributers

- Marietta Lee (hPolδ)
- Patricia Opresko (WRN)
- Thomas Kunkel (yPolδ,ε)
- Linda Bloom (RFC)
- Marila Cordeiro-Stone; G. Lucian Moldovan (cell lines)

Funding over the years:

Jake Gittlen Memorial Cancer Research Foundation
NIGMS/ R01 GM087472
NCI/ R01 CA100060
PA State Department of Health Tobacco Settlement Funds